

This is a repository copy of *Methods to compare Adverse Events in Twitter to FAERS, Drug Information Databases, and Systematic Reviews : Proof of Concept with Adalimumab*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/134658/>

Version: Published Version

---

**Article:**

Smith, Karen, Golder, Susan Pamela orcid.org/0000-0002-8987-5211, Sarker, Abeed et al. (3 more authors) (2018) *Methods to compare Adverse Events in Twitter to FAERS, Drug Information Databases, and Systematic Reviews : Proof of Concept with Adalimumab*. DRUG SAFETY. ISSN 0114-5916

<https://doi.org/10.1007/s40264-018-0707-6>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Methods to Compare Adverse Events in Twitter to FAERS, Drug Information Databases, and Systematic Reviews: Proof of Concept with Adalimumab

Karen Smith<sup>1</sup> · Su Golder<sup>2</sup> · Abeed Sarker<sup>3</sup>  · Yoon Loke<sup>4</sup> · Karen O'Connor<sup>3</sup> · Graciela Gonzalez-Hernandez<sup>3</sup>

© The Author(s) 2018

## Abstract

**Introduction** Adverse drug reactions (ADRs) are associated with significant health-related and financial burden, and multiple sources are currently utilized to actively discover them. Social media has been proposed as a potential resource for monitoring ADRs, but drug-specific analytical studies comparing social media with other sources are scarce.

**Objectives** Our objective was to develop methods to compare ADRs mentioned in social media with those in traditional sources: the US FDA Adverse Event Reporting System (FAERS), drug information databases (DIDs), and systematic reviews.

**Methods** A total of 10,188 tweets mentioning adalimumab collected between June 2014 and August 2016 were included. ADRs in the corpus were extracted semi-automatically and manually mapped to standardized concepts in the Unified Medical Language System. ADRs were grouped into 16 biologic categories for comparisons. Frequencies, relative frequencies, disproportionality analyses, and rank ordering were used as metrics.

**Results** There was moderate agreement between ADRs in social media and traditional sources. “Local and injection site reactions” was the top ADR in Twitter, DIDs, and systematic reviews by frequency, ranked frequency, and index ranking. The next highest ADR in Twitter—fatigue—ranked fifth and seventh in FAERS and DIDs.

**Conclusion** Social media posts often express mild and symptomatic ADRs, but rates are measured differently in scientific sources. ADRs in FAERS are reported as absolute numbers, in DIDs as percentages, and in systematic reviews as percentages, risk ratios, or other metrics, which makes comparisons challenging; however, overlap is substantial. Social media analysis facilitates open-ended investigation of patient perspectives and may reveal concepts (e.g. anxiety) not available in traditional sources.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s40264-018-0707-6>) contains supplementary material, which is available to authorized users.

---

✉ Graciela Gonzalez-Hernandez  
gragon@penmedicine.upenn.edu

<sup>1</sup> Rueckert-Hartman College for Health Professions, Regis University, Denver, CO, USA

<sup>2</sup> Department of Health Sciences, University of York, York, UK

<sup>3</sup> Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>4</sup> Norwich Medical School, University of East Anglia, Norwich, UK

## Key Points

Social media are a robust source of health-related data and may serve as complementary sources of adverse drug reaction (ADR) information from the patient perspective.

Analyses of social media posts allow for open, scientific investigation of ADRs that may be either not reported or underreported in spontaneous reporting systems and primary literature, thus contributing to a more complete safety profile.

Challenges prevent current natural language processing methods to automatically map all consumer ADR expressions to standard forms. Improvements to automatic text processing approaches should make the methods presented here scalable by reducing the annotation burden.

## 1 Introduction

Adverse drug reactions (ADRs) are a significant cause of morbidity and mortality worldwide, are responsible for approximately 5.3% of hospital admissions, and are estimated to rank between fourth and sixth in cause of mortality in the USA, making them crucial in healthcare decision making [1–4]. Recent studies have highlighted ADRs as costly and their discovery a public health priority [4–6]. The seriousness of the problem has led to investigations of novel methods to discover and assess ADR information from distinct sources [7–10]. The World Health Organization (WHO) defines an ADR as “a response to a drug which is noxious and unintended, and which occurs at doses normally used in man for the prophylaxis, diagnosis, or therapy of disease, or for the modifications of physiological function” [4]. The importance of ADRs should not be understated given that many drugs may result in harm and, when different options are available, avoidance of ADRs may be a deciding factor in formulary inclusion or treatment choice. There may also be narrow benefit-versus-harm trade-offs. A number of efforts originating from different parts of the globe attempt to systematically identify and report ADRs. However, comprehensive detection and reporting of ADRs remains incomplete [2, 4, 11]. Each reporting source has limitations. In particular, underreporting is apparent in traditional pharmacovigilance data, clinical trials, and other types of studies [12]. Drug information databases (DIDs), which are compilations of information primarily used by medical professionals, are limited by timeliness of updates from primary literature and package insert data. In recent years, social media have

emerged as promising sources of timely data that are currently underused and could help supplement data from other sources [13, 14]. Little is known about the similarities and differences between ADRs obtained from social media and traditional sources.

The primary objective of this study was to explore methods to elucidate the similarities and differences between data extracted from social media and other ADR reporting systems in an attempt to create a complete ADR profile for a single medication. We selected adalimumab as the focus of this study because it is approved for chronic diseases with significant health burdens, has been a top-selling drug in the USA for many years, and has both common and rare adverse events [15–18]. Adalimumab is a biologic, a class of medications that are large molecules manufactured in living cells for the treatment of diseases. It is a monoclonal antibody that binds to tumor necrosis factors, ultimately resulting in decreased inflammatory activity in immune-based conditions. Its primary use is in immune-based arthritis, Crohn’s disease, ulcerative colitis, and psoriatic arthritis, although additional indications have been approved by the US FDA. Adalimumab has many common side effects such as injection site reactions, and, rarely, it increases the risk of serious infections [17]. We hypothesized that social media provide information to support what is known about medications, along with additional information that is not available from traditional systems.

### 1.1 Challenges with Adverse Drug Reaction (ADR) Reporting

Precise ADR reporting rates are unknown but estimated at approximately 10–20% [12, 19, 20]. Data collection and reporting varies in different settings. For example, ADRs measured in a hospital setting have been estimated at 86%, whereas the rate in outpatient settings may be 16–48% [21–25]. Further differences may be seen in both observational and experimental studies, all of which suffer from underreporting [12]. Problems arise with differences in reporting rates, imprecise evaluation and estimation of ADRs, and a scarcity of recent studies to describe ADR rates [3, 26–29]. The following subsections detail some of these limitations.

#### 1.1.1 ADRs Reported in Primary Literature

DIDs and systematic reviews rely on primary literature, such as reports of randomized controlled trials (RCTs). In general, RCTs are designed and powered to explore efficacy and often are neither large enough nor have sufficient follow-up to identify rare, long-term ADRs or ADRs that occur after the drug has been discontinued. RCT data may be limited if trials exclude specific patient populations such as children,

elderly, pregnant women, patients with multiple diseases, and those with potential drug interactions. Aggregating data from RCTs is another method to examine the association between medications and ADRs. Systematic reviews aim to identify, evaluate, and summarize findings of relevant studies, mostly RCTs. When appropriately conducted, they provide reliable estimates about the effects (beneficial and adverse) of interventions.

Detection of ADRs through observational studies can be problematic because they are typically not designed with ADR detection as the primary outcome. Pharmacoepidemiologic (PE) studies measure the effects and ADRs of drugs in large populations. PE studies may use secondary data sources such as administrative health claims databases [30]. Health claims are a robust source of real-world data, yet the databases were not designed to detect ADRs. These observational studies may be valuable for hypotheses generation, education applications, and even pharmacovigilance, yet it is difficult to establish cause and effect relationships, the opportunity for bias is great, and they rank low in evidence-based medicine. More common are case series and case reports, which rely on researchers or motivated clinicians to investigate, evaluate, and report the case(s) [31].

### 1.1.2 ADRs and Spontaneous Reporting Systems

The FDA Adverse Event Reporting System (FAERS) is a spontaneous reporting mechanism in the USA that has mandatory and voluntary components. FAERS may be one of the most efficient current methods to capture rare events that are associated with drug use [32]. FDA requires manufacturers to report serious ADRs within 15 days of receipt. ADR reports are maintained in the FAERS database, and manufacturers are estimated to provide 80% of pharmaceutical reports [33]. Reporting by consumers and healthcare professionals accounts for the remaining reports, and the level of detail and content in each of these reports may limit their usefulness. Healthcare professional and consumer reporting is voluntary, leading to both bias in what is reported and underreporting. Healthcare professionals may be unsure of who is responsible for submitting an ADR report (20–36%) [11]. Further obstacles to healthcare professionals reporting are insufficient time to report and unclear reporting processes; they may also be unsure of the specific drug causing the reaction.

Other organizations maintain ADR reporting databases, for example, the Institute for Safe Medication Practices, but its focus is safe use and error prevention [34]. The Agency for Healthcare Research and Quality supports research to evaluate ADRs but does not compile them in a database [35]. The American Society of Health System Pharmacists has a recommended plan for reporting ADRs, but the focus is primarily hospital based [36]. The WHO and European

countries have spontaneous surveillance mechanisms, such as the UK Medicines and Healthcare Products Regulatory Agency's Yellow Card program, the European Medicines Agency's EudraVigilance, and the WHO Uppsala Monitoring Center in Sweden, all of which have similar reporting challenges.

### 1.1.3 ADRs Reported in Social Media

Social media platforms have been explored in the recent past as a potential resource for pharmacovigilance [13]. Given the limitations of spontaneous reporting systems (SRS) and other sources, social media in general and Twitter specifically was identified as a source to quantify adverse events [13, 14, 37]. Social media has the potential to inform and augment adverse event reporting systems and self-reported perception of health that may not have been previously collected from other sources. Patient reporting brings novel information, more detail, and information on severity and impact of ADRs in daily life [38]. Furthermore, ADRs associated with over-the-counter medications may not be captured in either hospital or ambulatory settings but appear in social media.

In prior research, we developed natural language processing (NLP) methods that address specific challenges of mining health-related information from social media texts [13, 39, 40]. The properties of social media texts that pose NLP challenges include misspellings, data imbalance, non-standard expressions, and noise, to name a few. For social media-based pharmacovigilance, we completed large-scale efforts of manual annotation and used them to develop supervised classification methods to filter out noise [41] and information extraction methods to extract standard and non-standard mentions of ADRs [39]. These advances mainly allow the reduction of the volume of data that needs to be manually annotated, given that any study on specific ADRs or their frequencies to estimate incidence or detect a signal [such as proportional reporting ratio (PRR)] requires mapping of the extracted mentions to standard ADR nomenclature [such as the Unified Medical Language System (UMLS) or MedDRA], a task commonly known as normalization. This remains an unresolved challenge, with lexicon-based approaches (e.g. MetaMap) performing very unreliably [42, 43] and missing all but the almost exact mentions of the ADRs. Effective automatic normalization methods will potentially enable large-scale comparisons of many drug-ADR pairs from social media to other sources. When this study was conducted, because an effective normalization approach was lacking, we chose to use automatic methods for ADR extraction only, followed by manual mapping of the mentions to standard nomenclature. Manual normalization of all posts reporting an ADR removed the potential bias of only using posts that could be mapped automatically.

We compared ADRs from Twitter to known sources such as DIDs, systematic reviews, and FAERS to determine similarities and differences. This information may reflect patient perceptions or indicate signals of medication-related problems. The annotation effort plus the three-way comparison of social media data to other sources forced us to limit the study to the ADRs associated with a single drug. While this manual effort was cumbersome, it was necessary to ensure accurate comparisons, and it lays down the framework for future larger-scale comparisons of multiple drugs and ADRs from different sources. We hypothesize that social media provide not only information about medication use that supports what is known about medications but also useful information that is not available from traditional adverse event reporting systems.

## 2 Methods

Twitter posts were collected from June 2014 to August 2016 via the public application programming interface (API) (<https://dev.twitter.com/streaming/public>) using the keywords Humira and adalimumab and their automatically generated misspellings [40]. Tweets were processed by ADRMine [39], which is a sequence-labeling system relying on conditional random fields, to extract ADRs. All extracted ADRs were manually annotated as perceived ADRs or some other categories. In addition, to assess what the automatic system could be missing, a random sample of 1000 tweets were selected from those not identified by ADRMine and similarly annotated. All manual annotations were performed following the guidelines described by O'Connor et al. [44]. Identified ADRs were mapped to UMLS [45] concept IDs.

Twitter adalimumab ADRs were aggregated into broad categories of biologic systems to facilitate comparison between the different ADR sources. For example, local injection site reactions are known and common for adalimumab and may be identified with the UMLS concept “injection site burning” or “injection site bleeding” or “bruising,” and others. FAERS reports these local reactions as “injection site pain” or “injection site hemorrhage” and others, whereas DIDs report the reactions in one subcategory of local injection site ADRs. For this study, these local dermatologic reactions were combined into one category, and a similar process was used to create each biologic system category.

We conducted disproportionality analyses by computing PRRs, which is a measure of disproportionality in signal detection [46, 47]. This measure has been used for SRS, and the score was customized to suit social media data [48]. The goal of disproportionality analysis is to detect drug–ADR pairs that are reported more frequently than other pairs of concepts. Relatively small numbers of reports may lead to identifiable signals. Table 1 presents the contingency matrix

**Table 1** Disproportionality two-by-two contingency matrix

	User posts with the suspected ADR	User posts without the ADR
User posts mentioning adalimumab	A	B
All other posts	C	D

ADR adverse drug reaction

for the disproportionality measure, which is given by the following equation:

$$PRR = \frac{A/(A+B)}{C/(C+D)}.$$

### 2.1 Comparison of Tweets and Known Sources of ADR Reporting

We compared ADR categories mentioned in tweets with three known sources of ADR reporting: FAERS, DIDs, and systematic reviews of adalimumab.

#### 2.1.1 Comparison Metrics: Frequency, Ranking, Relative Frequency of ADRs

Frequencies were compared and ranked as the absolute percentages identified across sources. To compare the relative magnitude of differences between the ADR categories, we computed the relative frequencies of the most mentioned categories of ADRs. “Pain” was defined as the index comparator with a value of 1.0 because it was reported similarly across Twitter, FAERS, and the DIDs. To obtain a relative frequency of an ADR compared with pain, the percentage reporting that ADR was divided by the percentage reporting pain. For example, 17.2% of tweets mentioned musculoskeletal complaints, and 10% of tweets mentioned pain. The relative frequency of musculoskeletal mentions would be 1.7% in FAERS.

#### 2.1.2 US Food and Drug Administration Adverse Drug Event System

We obtained a report of ADRs for adalimumab using Opened API (ResearchAE.com) from 1 June 2014 through August 2016. ADR categories were compared between tweets and FAERS reports by frequency, rank order, relative frequencies, and PRR metrics.

#### 2.1.3 Drug Information Databases

We compared the frequency of ADR tweets with similar events mentioned in the three major DIDs utilized by



healthcare professionals—Micromedex<sup>®</sup>, Lexicomp<sup>®</sup>, and Clinical Pharmacology<sup>®</sup>—using a composite frequency [see Electronic Supplementary Material (ESM) 1]. These databases are common sources of drug information for clinicians and use multiple sources, including primary literature, to present evidence-based efficacy and adverse event data. ADR categories were compared between tweets and DIDs by frequency, rank order, relative frequencies, and PRR metrics.

### 2.1.4 Systematic Reviews of Adalimumab

We conducted an overview of systematic reviews of adalimumab to identify ADRs and compared the ADRs identified in systematic reviews of adalimumab with the frequencies of the ADRs in the tweets. Systematic reviews are considered the gold standard in evidence-based medicine. They are often used as the basis of clinical guidelines and are considered as guidance in evidence-based policy decisions. To identify ADRs associated with adalimumab in systematic reviews, a range of databases including Epistemonikos (<https://www.epistemonikos.org/>) and the DARE archive (<https://www.crd.york.ac.uk/CRDWeb/ResultsPage.asp>) were searched in February 2017 for synonyms of the terms Humira and adalimumab. No date or language restrictions were applied to the searches. Systematic reviews were included if Humira was one of the primary interventions and investigators had searched for and presented usable data related to the ADRs of this intervention.

Comparing data collected from social media with outputs from systematic reviews was challenging. We undertook a two-step approach where we first presented findings and categories of ADRs identified via Twitter and systematic reviews. The second step involved analyses to examine the strength of the relationship between adverse events and adalimumab—a rank order of the ADR frequencies. This gave an indication of agreement between the most frequently occurring ADRs from each source. These approaches used the data or statistics as presented by the systematic review authors, which in most instances meant we limited our comparison to the absolute numbers from the active arm only. However, this was more comparable to the data collected from tweets, where a control arm was not available.

## 3 Results

A total of 10,188 tweets were collected and 2617 potential ADRs identified automatically by ADRMine. ADRMine obtained an F-measure of 0.58 (recall 0.91, precision 0.43) over the dataset. The high recall obtained by the system suggests that most true ADR expressions were captured by the system. Manual review of these and an additional 1000 tweets randomly chosen from those not identified by

ADRMine resulted in 801 true ADRs, which were mapped to 232 unique UMLS concept IDs. Among others, 112 were ambiguous because it was unclear whether they were ADRs, 259 were ambiguous because the mention may have referred to either the indication for the drug or the ADR, 250 referred to the indication for adalimumab, 98 referred to an ADR or indication for another drug, six referred to a beneficial effect of the drug, 32 were duplicates, and 1162 mentioned the drug but not an ADR or indication for the drug or other classifiable mention. These tweets were excluded from the analysis. Table 2 presents sample Twitter posts and type categorization.

### 3.1 Classification of Tweets into Categories Based on UMLS Concept IDs

After excluding tweets that did not mention an ADR, extracted UMLS concept names describing similar adverse events were combined into one biologic category, where possible, to measure the frequency of a broad concept as previously described. For example, UMLS concept names, including “knee pain,” “muscles pain,” and multiple other “pain” mentions were grouped into one general “pain” category. Other categories created by pooling related concepts included “abdominal issues,” “allergic reactions,” “anxiety and mood mentions,” “dizziness and neurologic symptoms,” “fatigue symptoms,” “headaches,” “infections,” “cardiac symptoms,” and “sick mentions”. Finally, serious illnesses such as “Guillain Barre,” “myocardial infarction,” and “fatal outcome” were noted. Table 3 presents the top 16 aggregated categories mentioned in adalimumab tweets, along with the DID and FAERS report frequencies, frequency ranks, and relative index values.

### 3.2 Comparison of ADRs in Tweets with FAERS and Drug Information Databases

Absolute frequency ranking and relative ranking with index values are presented in Table 3. Local “injection site reactions,” “fatigue,” and “pain” accounted for the majority of concept categories for Twitter mentions at 187 (23.7%), 136 (17.2%), and 79 (10.0%), respectively; this was similar to FAERS reports, at 24.4, 11.4, and 11.6%, respectively. While these ADRs were frequently reported in the drug information databases, the top three events were “injection site reactions,” “skin/dermatologic reactions,” and “headache”. PRR scores for the term “pain” were high at 16.4 when the scores for the specific mentions (e.g. muscle pain and chest pain) were combined. The PRR score most applicable to “injection site pain” was “welts,” which also had a high score, of 8.4.

The relative trend of the top ADRs in Twitter of “injection site pain,” “musculoskeletal fatigue,” “gastrointestinal complaints,” and “neurologic complaints”, including “anxiety”,

**Table 2** Examples of tweets expressing adverse drug reactions, indications, beneficial effects, and other information. Unified medical language system concept codes are shown in the first column

UMLS concept code	Comment type	Comment
Pain	ADR	Humira is the worst pain I have ever felt—and I've felt some pretty bad pain in life
Pain	Unsure	@username can't sleep in pain, reminds me of your super woman strength to cope with it! How did ur first Humira injection go babe?
Pain	Indication or ADR	I just gave myself my first Humira injection. Ouch! Any suggestions on reducing the pain and making it more tolerable?
Pain	Other drug	I had terrible joint pain with Remicade. I switched to Humira and have had no major side effects
Pain	Beneficial	I'd like to add, Humira is kicking RA's ass! I've got energy! Pain & swelling down a lot! Haven't taken ibuprofen in 2 days either :)
Fatigue	ADR	Humira made me perpetually fatigued and sick
Fatigue	Unsure	One of those days—wicked tired and I have to get going to make it to physical therapy today. Two days after each Humira shot is sucky
Fatigue	Other drug	I think I am still trying to get over the tiredness from Ibiza and Humira is due
Fatigue	Beneficial	I could swear that Humira is giving me so much energy

ADR adverse drug reaction, UMLS unified medical language system

**Table 3** Adverse event categories mentioned in adalimumab tweets

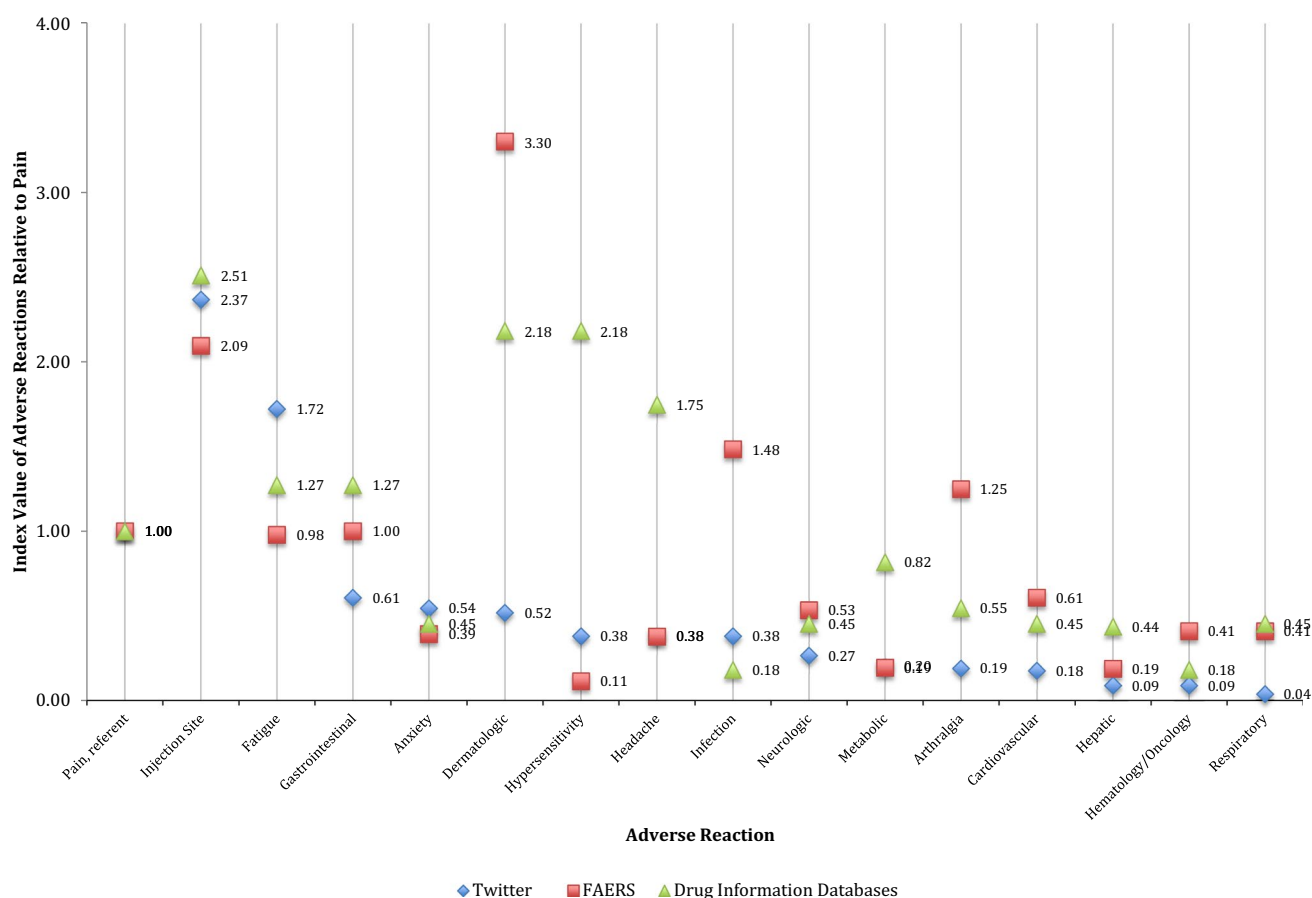
Adverse Event Category	Twitter				FAERS				Drug Information Databases		
	n	Percent	Frequency Rank	Index value <sup>a</sup>	n	Percent	Frequency Rank	Index Value <sup>a</sup>	Percent	Frequency Rank	Index Value <sup>a</sup>
Local: Injection Site	187	23.70	1	2.4	603	24.36	2	2.1	13.80	1	2.5
Musculoskeletal: Fatigue/ Weakness/ Spasms/ Malaise	136	17.24	2	1.7	282	11.39	7	1.0	7.00	5	1.3
Pain,all	79	10.01	3	1.0	288	11.64	5	1.0	5.50	7	1.0
Gastrointestinal	48	6.08	4	0.6	288	11.64	5	1.0	7.00	5	1.3
Neurologic: Anxiety/ Depression/ Insomnia/ Panic/ Mood	43	5.45	5	0.5	113	4.57	12	0.4	2.50	10	0.5
Dermatologic	41	5.20	6	0.5	950	38.38	1	3.3	12.00	2	2.2
Neurologic: Headache	30	3.80	7	0.4	109	4.40	13	0.1	12.00	2	2.2
Infection	30	3.80	7	0.4	427	17.25	3	0.4	9.60	4	1.7
Hypersensitivity	30	3.80	7	0.4	33	1.33	16	1.5	1.00	15	0.2
Neurologic: CNS Dizzy-fall/ Gait/ Groggy/ Memory/ Confusion	21	2.66	10	0.3	154	6.22	9	0.5	2.50	10	0.5
Endocrine/ Metabolic	16	2.03	11	0.2	56	2.26	14	0.2	4.50	8	0.8
Musculoskeletal: Arthralgia/ Joint/ Bone complaints	15	1.90	12	0.2	360	14.55	4	1.3	3.00	9	0.5
Cardiovascular	14	1.77	13	0.2	175	7.07	8	0.6	2.50	10	0.5
Hepatic	7	0.89	14	0.1	54	2.18	15	0.2	2.40	14	0.4
Hematology/ Oncology	7	0.89	14	0.1	118	4.77	10	0.4	1.00	15	0.2

CNS central nervous system, FAERS US FDA Adverse Event Reporting System

<sup>a</sup>Pain referent

“depression”, and “mood” are similar, as illustrated in Fig. 1. Not all relative reports were similar. FAERS reported dermatologic ADRs three times more frequently than pain (index value 3.3), whereas Twitter reports of dermatologic ADRs were 0.5 times that of pain. The DID reports of dermatologic ADRs were more similar to those in FAERS than on

Twitter, with an index value of 2.2. The five ADR categories with notable disparity between the three sources were “dermatologic”, “hypersensitivity”, “headache”, “infection”, and “joint and bone” ADRs. Of those categories, “dermatologic”, “hypersensitivity”, and “bone and joint ADRs” had the highest index values when compared with pain in the



**Fig. 1** Ranking of index values for adverse drug events by data source. *FAERS* US FDA Adverse Event Reporting System

FAERS reports (3.3, 1.5, 1.3, respectively). “Hypersensitivity”, “infection”, and “headache” ADRs had the highest index value relative to “pain” in the drug information databases (2.2, 2.2, and 1.7, respectively). The relative ranking of index values for Twitter were less than the other data sources in “gastrointestinal”, “joint and bone”, “cardiovascular”, “hepatic”, “hematology and oncology”, and “respiratory” categories (Fig. 1).

### 3.3 ADRs Identified Through Systematic Reviews

Database searches identified 38 systematic reviews evaluating adalimumab. After assessment of each paper, we included 20 systematic reviews (see ESM 2). Seven reviews were excluded because they did not evaluate any adverse effects [49–55], three were excluded because they pooled serious ADRs or all ADRs without naming or quantifying the adverse effects [56–59], and another was excluded because it evaluated pharmacoeconomic studies [60]. Five systematic reviews contained data on combination therapy rather than adalimumab monotherapy or compared

adalimumab therapy with another therapy [61–65]. One review [66] contained only two RCTs, both of which were already included in another review [67] with the same data extracted and was therefore excluded. If a study was included in more than one review, we checked to see whether different outcomes were measured and removed duplicates. No further systematic reviews were excluded on this basis. In total, 20 systematic reviews remained for inclusion to examine ADRs to adalimumab (Appendix 2 in the Supplementary material). One included study [15] appeared to have been updated [68], but the included studies were not listed. In this instance, we extracted data from both papers but did not include both sets of data in our results.

Adalimumab was evaluated for different indications, study designs, and results measures in the 20 included systematic reviews. Indications included rheumatoid arthritis, inflammatory arthritis, ankylosing spondylitis, psoriasis, ulcerative colitis, inflammatory bowel disease, or Crohn’s disease. While half of the reviews were limited to RCTs only [67, 69–77], eight included RCTs and other study designs (such as case series or non-randomized trials) [15, 68,



78–83], and two included case reports, case series, cohort studies, or uncontrolled trials, not RCTs [84, 85]. Results from the reviews were either reported as rates of adverse effects (nine reviews) [67, 70, 71, 73, 76, 79, 80, 83, 84], events per patient-years (six reviews) [15, 68, 69, 72, 85, 86], risk ratios (one review) [77], or were simply listed with little or no numerical data (four reviews) [75, 78, 81, 82].

We were only able to obtain the rank order of adverse effects from ten systematic reviews [67, 68, 70, 73, 74, 76, 77, 80, 83, 84]. Two of the nine systematic reviews reporting rates of ADRs [71, 79] and three of the six systematic reviews reporting events per patient-years [15, 72, 85] did not provide numerical data for all the adverse effects listed, and one systematic review only evaluated one ADR (cancer) [69]. For the systematic reviews that did not provide adequate numerical data, we only compared whether particular ADRs were mentioned [15, 71, 72, 75, 77–79, 81, 82, 85].

### 3.3.1 Comparison of Mentioned and Ranked ADRs

Within the ten systematic reviews for which we were able to rank order ADRs, the two categories from the top 16 categories in Twitter not covered in these systematic reviews were “neurologic: anxiety/depression/insomnia/panic/mood” and “endocrine/metabolic”, which were the 5th and 11th top adverse effects on Twitter, respectively. To compare the ranking of ADR categories in systematic reviews to Twitter, we first compared the overall results from the ten reviews with rank order data and then carried out a more detailed analysis with the included RCTs from these reviews that reported rates in the treatment and placebo arms. No single adverse event category was included in all ten systematic reviews; however, “infection” ranked seventh on Twitter, was covered in nine systematic reviews, and ranked first overall in the systematic reviews. “Injection site reactions” (top ADR on Twitter) and “hematology/oncology” (14th on Twitter) were listed in six reviews. “Injection site reactions”

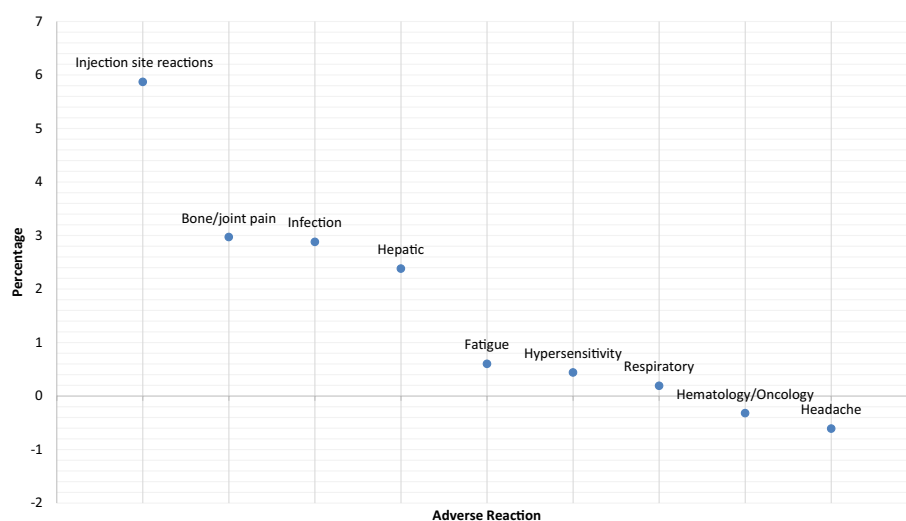
were ranked second overall, and “hematology/oncology” was ranked fifth overall in the systematic reviews. Interestingly “infection” and “hematology/oncology” were much more prominent ADRs in the systematic reviews than in social media. Conversely, “musculoskeletal: fatigue” (third on Twitter) and “pain” (fourth on Twitter) were much more prominent in social media than in systematic reviews (mentioned in one and four reviews, respectively). Some ADRs reported in the systematic reviews were not in the top 16 categories from Twitter. Most notable were serious ADRs such as “death”—although this was not common.

Figure 2 displays the results of a more detailed analysis using rates of ADRs reported in the systematic reviews [70, 73, 76, 83]. This figure was produced by summing the adverse events from the RCTs in each category to calculate the absolute percentage difference for adalimumab versus placebo, and it displays the rank order of attributable frequency, which is the intervention ADR rate minus the control event rate. Only “infection” had complete data in all 11 RCTs and “injection site” was reported in nine RCTs. One major issue was that “pain” was not reported as an ADR in any of the trials, primarily because pain is non-specific and could be applicable to any biologic system or the disease. Figure 2 highlights that the investigators may only be interested in measuring infections and local reactions, while important ADRs remain poorly ascertained.

## 4 Discussion

This study illustrates the similarities and differences in ADRs discovered for adalimumab from different sources and highlights the difficulties of comparing or combining data from the different sources because each has a unique set of limitations. FAERS is a voluntary reporting system and, while the total number of ADRs reported is known, the total number of individuals exposed to the drug are not.

**Fig. 2** Change in risk for adverse drug reactions in those receiving adalimumab versus placebo in randomized controlled trials



Thus, incidence cannot be determined from FAERS data. Incidence may be measured in clinical studies, which are the primary source of information for drug information databases and systematic reviews. However, these studies may of short duration and have limited patient populations, and they are normally designed to detect efficacy. Most systematic reviews are also not designed with ADRs as the primary outcome; in many cases, the individual studies that contribute to the reviews report ADRs by percentage, incidence rates utilizing person time, or other crude measures. Given the limitations of these sources, it is reasonable to utilize social media as additional sources of patient-reported ADRs. We sought to verify the extent to which social media encapsulated knowledge regarding ADRs identified in traditional sources and to determine what additional information about ADRs social media data may provide us.

#### 4.1 Comparison with Related Work

Social media are a recognized source of ADR information within the research community, yet methods of extracting, analyzing, and interpreting it are many, as are the proposed uses [8, 13, 14, 26, 87–90]. Furthermore, automated approaches relying on social media show varying performances, and continued developments to improve precision are needed. Sarker et al. [13] conducted a review of studies that described automatic data mining approaches for ADR detection. They identified 22 studies that used health-related and general social media sites as the source for mining. The authors concluded that while health-related sources contained more specific ADR data, publicly annotated data to allow for further development of methods to identify ADR–drug pairs were lacking. Lardon et al. [87] conducted a review of ADR extraction from social media to determine what methods had been used to identify post-marketing pharmacovigilance data and evaluate the signals contained in user postings. The authors identified studies that focused on extraction and evaluation of ADR–drug pairs and found that four compared the data with those in FAERS, seven studies used experts to evaluate ADR comments, and none compared frequency and relative ranking of comments with multiple sources of pharmacovigilance data. We found that, when comparing ADRs in Twitter with those in FAERS, more common ADRs had similar relative values across sources, and likewise for the less common ADRs. Twitter ADRs that occurred at a moderate frequency were relatively underreported when compared with those in FAERS. For example, “dermatologic” ADRs had a relative value of 0.5 in Twitter and 3.3 in FAERS or 2.2 in the drug information databases (Fig. 1). “Hypersensitivity” was reported more frequently in the DIDs (index 2.2) and relatively less in Twitter and FAERS (0.4 and 0.1, respectively).

Our study utilized multiple sources, whereas a previous comparison only examined systematic reviews and social media [14]. Only 12 of the 51 identified studies used medical dictionaries in their search strategy to identify ADRs. There was general agreement between the extraction source and existing ADR data: over 80% of the ADRs found in Twitter were supported by the findings from other sources. Milder ADRs were reported at a higher frequency in social media. By comparison, we found that when evaluating the relative ranking of the ADR categories, the most frequently and infrequently experienced ADRs were similar across sources, with the exception of “dermatologic” ADRs reported in FAERS. The moderately reported ADRs were more likely to differ between Twitter, FAERS, and DIDs. Our finding that dermatologic ADRs were more often reported in FAERS is similar to previous findings that some social media ADRs are underrepresented compared with in pharmacovigilance systems [14]. The relative ranking of lesser-reported ADRs was similar across the sources. For example, cardiovascular, hepatic, and respiratory ADRs had relatively low reports in all the sources (Fig. 1).

A study similar to ours attempted to compare ADRs across sources (not social media) using an index value for relative comparisons from WHO reports, published case reports, and results of their meta-analysis [10]. The authors selected ADRs associated with amiodarone, a medication with known pulmonary, thyroid, and rare ophthalmic complications. There was little agreement within the sources for the top ADR, as cardiac problems were ranked highest in the authors’ meta-analysis but lowest in WHO and case reports. Thyroid problems were reported in the top three of each data source, indicating some level of agreement. Additionally, ophthalmic ADRs were reported to be rare, resulting in the low ranking in each of the sources. Our study, unlike that of Loke et al. [10], included social media as a source, and we found that the moderately reported ADRs varied in relative reporting frequency (dermatologic, hypersensitivity, headache, and infection).

One method employed to examine the potential of social media as an early warning system identified safety signals reported to the FDA then retrospectively determined whether the signals were present in Twitter and Facebook [91]. The authors selected ten drug–event pairs from FAERS then examined user posts to determine whether the event was mentioned before the FAERS signal. Their semi-automated analysis identified 13 posts in which the drug may have caused the event in question—six were identified as definitely causal, probable, or possible (46%). While they sought to determine whether the known drug–event pair signals were present in social media before FAERS reporting, our study examined the complete stream of postings coming anew for a specific drug without the advantage of hindsight. With the state of current automatic NLP methods,

it is relatively easier to look for single, expected effects than to find and analyze all mentions of any type of effect.

These studies underscore the challenges in comparing ADRs between sources. ADRs may be reported as a percentage of occurrence in exposed individuals in scientific literature, but difficulties arise because social media studies do not have a control treatment. Absolute measures of frequency may be the most straightforward comparison, but if a study reports a low incidence (e.g. 3 per 1000 patient-years), comparisons are difficult to describe. Relative measures are especially challenging, depending upon the comparator and control group. Control groups may differ between trials, so the relative effect also differs. Further complicating comparisons is concomitant therapy (e.g. adalimumab plus methotrexate). Methotrexate may be adjusted or dose titrated in a trial, so the adalimumab-treated patients may be receiving a different dose of the concomitant medication. This is problematic when ADRs are related to dose response.

DIDs compile ADRs from multiple sources, including scientific literature, and are therefore subject to the same frequency challenges. The FAERS database of post-marketing ADRs is subject to additional challenges because it can only report absolute occurrence numbers. Even with multiple reporting systems, ADRs are underreported, and social media appear to be sources to augment existing reporting of ADRs and health perceptions [87, 92, 93].

In addition to issues with comparing ADRs with different measures, DIDs categorize and display ADRs differently, which may lead to interpretation differences. Micromedex lists ADRs in the “Quick Answers” section as either common or serious. Categories within this section include anatomical systems arranged by frequency of occurrence. For example, “dermatologic” includes injection site reaction of 5–20%, depending on the age category of the user. The “In Depth Answers” section lists eight subcategories within the dermatologic section. Lexicomp presents adverse reactions by frequency in a similar fashion to Micromedex, with three categories of severity (<1%, 1–10%, >10%). An anatomical system may appear in each of these categories. It is difficult to know precisely what to expect in terms of ADRs or how to compare them with other sources.

When compared with Twitter, systematic reviews have the advantage of including studies that have a control group. However, many of the adverse events listed in the treatment arm have similar frequencies in the control or placebo arm. Computation of an odds ratio for the ADRs could solve the difficulty in determining differences, but this is typically not feasible as most studies are only powered to detect differences in the primary outcome. We used the raw data in the treatment group because this is more easily comparable with the data collected from social media. Some systematic reviews simply listed ADRs that occurred in the included studies with no frequencies or gave frequencies for selected

ADRs only. The majority of systematic reviews focused on clinical effectiveness. Outcomes in systematic reviews should be pre-specified at the protocol stage, at which time the ADRs may be unknown. These factors may explain some of the poor reporting of the ADR outcomes and may give an impression that there are no significant differences in adverse events. The various indications and range of study designs included in the systematic reviews meant that a plot of relative ADR frequencies from the systematic reviews would be likely to produce a large scatter because they were all so different. This confounds our subsequent comparison with other sources. Therefore, we could not produce a single consistent estimate from all the relevant systematic reviews that could be used to compare against Twitter or the DIDs, or FAERS data. Hence, we produced a rank order from only a small proportion of the systematic reviews available.

A strength of our study is that we identified ADRs that may not be known, were not reported to the FDA, or were not well described elsewhere. For example, UMLS concepts in tweets that were not mentioned in the DIDs were “sleep” and “nervous”. “Sleep” tweets included both ends of the spectrum, such as “day one on the new Humira and I sleep the entire morning away”, and “this Humira keeps me either awake or it only lets me sleep for 2 h”. In general, however, Twitter users appeared to associate lack of sleep with the drug, whereas the majority of tweets containing the ADR “nervousness/anxiety” referred not to how the drug made them feel but to the anxiety caused by the act of self-injecting. For example, “still get nervous when injecting sometimes” or “always get nervous giving myself a Humira shot”. These tweets occur at a different frequency than ADRs reported to the FDA, presumably because they are not directly a cause of the drug but are related to patient feelings about the drug. These two examples support our hypothesis that ADRs and patient perspectives (known or unknown) are available in social media. The unknown concepts here may reflect a true ADR or the subjective perceptions or sentiments associated with the drug but, in either case, require further study.

Current ADR identification efforts have benefits and challenges. Challenges in interpreting tweets include the limited length and the reflection of patients’ perceptions of a condition or drug when the condition may or may not be associated with the drug at all. Furthermore, the data mined may have privacy concerns. Additional limitations include the reliability of tweets, but the semi-automatic process does provide a mechanism for excluding unrelated comments while capturing user perceptions. Scalability may be considered as limiting because the combined automated process of extracting tweets and subsequently compiling the DID reporting and conducting the systematic review took an enormous amount of work. However, limitations of automatic approaches may be alleviated as further advances are

made in NLP methods applied to social media. Ongoing advances in NLP and normalization techniques will allow this type of comparison to be completed for multiple drugs. Studies such as ours motivate and give significance to those efforts.

## 5 Conclusion

Twitter is a robust source of health-related data, and therefore it is important to continue to refine methods to best utilize it. We conducted a prodigious comparison of tweets to known sources of ADR information and determined the level of agreement. Generally, concepts were in moderate agreement with known ADRs, and concepts were found in Twitter that were not in DIDs (e.g. nervous/sleep). This study demonstrates that it is possible and worthy to harvest and compare ADRs found in social media with those from typical sources to augment what is known, while a large-scale effort (that includes multiple medications) is still difficult given the level of effort required. Challenges to analyzing social media may be attenuated as further advances are made to automatic NLP methods.

**Acknowledgements** The authors thank the anonymous reviewers for their helpful feedback.

## Compliance with Ethical Standards

**Funding** This work was supported by National Institutes of Health (NIH) National Library of Medicine (NLM) grant number NIH NLM 5R01LM011176. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NLM or NIH.

**Conflict of interest** Karen Smith, Su Golder, Abeed Sarker, Yoon Loke, Karen O'Connor, and Graciela Gonzalez-Hernandez have no conflicts of interest that are directly relevant to the contents of this study. After the commencement of this study, Graciela Gonzalez-Hernandez received funding from AbbVie Inc. for an unrelated research project focusing on using social media to complement pregnancy registry information. AbbVie or its employees had no influence on or knowledge of the design, methodology, or findings discussed in this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Kongkaew C, Noyce PR, Ashcroft DM. Hospital admissions associated with adverse drug reactions: a systematic review of prospective observational studies. *Ann Pharmacother*. 2008;42(7):1017–25.
- FDA Adverse Event Reporting System (FAERS). <http://www.fda.gov/Drugs/InformationOnDrugs/ucm135151.htm>. Accessed 6 Jan 2016.
- Heinrich J. Adverse drug events: Substantial problem but magnitude uncertain: United States General Accounting Office; 2000. p. 6. <http://www.gao.gov/new.items/he00053t.pdf>. Accessed 10 Aug 2018.
- World Health Organization. Safety of medicines—a guide to detecting and reporting adverse drug reactions—why health professionals need to take action: World Health Organization; 2017. <http://apps.who.int/medicinedocs/en/d/Jh2992e/3.html>. Accessed Apr 2017.
- Lahue BJ, Pyenson B, Iwasaki K, Blumen HE, Forray S, Rothschild JM. National burden of preventable adverse drug events associated with inpatient injectable medications: healthcare and medical professional liability costs. *Am Health Drug Benefits*. 2012;5(7):1–10.
- Bourgeois FT, Shannon MW, Valim C, Mandl KD. Adverse drug events in the outpatient setting: an 11-year national analysis. *Pharmacoepidemiol Drug Saf*. 2010;19(9):901–10.
- Harpaz R, DuMouchel W, LePendur P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin Pharmacol Ther*. 2013;93(6):539–46.
- Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf*. 2014;37(5):343–50.
- Brown JS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf*. 2007;16(12):1275–84.
- Loke YK, Derry S, Aronson JK. A comparison of three different sources of data in assessing the frequencies of adverse reactions to amiodarone. *Br J Clin Pharmacol*. 2004;57(5):616–21.
- Stergiopoulos S, Brown CA, Felix T, Grampp G, Getz KA. A survey of adverse event reporting practices among US healthcare professionals. *Drug Saf*. 2016;39(11):1117–27.
- Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. *Drug Saf*. 2006;29(5):385–96.
- Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform*. 2015;54:202–12.
- Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *Br J Clin Pharmacol*. 2015;80(4):878–88.
- Burmester GR, Mease P, Dijkmans BA, Gordon K, Lovell D, Panaccione R, et al. Adalimumab safety and mortality rates from global clinical trials of six immune-mediated inflammatory diseases. *Ann Rheum Dis*. 2009;68(12):1863–9.
- Scheinfeld N. Adalimumab: a review of side effects. *Expert Opin Drug Saf*. 2005;4(4):637–41.
- Leombruno JP, Einarson TR, Keystone EC. The safety of anti-tumour necrosis factor treatments in rheumatoid arthritis: meta and exposure-adjusted pooled analyses of serious adverse events. *Ann Rheum Dis*. 2009;68(7):1136–45.
- Phillippidis A. Top 15 best selling drugs of 2016. Genetic engineering and biotechnology news. 2017. <https://www.genengnews.com/the-lists/the-top-15-best-selling-drugs-of-2016/77900868>. Accessed 15 Dec 2017.
- Leape LL. Reporting of adverse events. *N Engl J Med*. 2002;347(20):1633–8.
- Pillans PI. Clinical perspectives in drug safety and adverse drug reactions. *Expert Rev Clin Pharmacol*. 2008;1(5):695–705.



21. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *JAMA*. 1997;277(4):307–11.
22. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annet JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA*. 2006;296(15):1858–66.
23. Cohen AL, Budnitz DS, Weidenbach KN, Jernigan DB, Schroeder TJ, Shehab N, et al. National surveillance of emergency department visits for outpatient adverse drug events in children and adolescents. *J Pediatr*. 2008;152(3):416–21.
24. Jha AK, Kuperman GJ, Rittenberg E, Teich JM, Bates DW. Identifying hospital admissions due to adverse drug events using a computer-based monitor. *Pharmacoepidemiol Drug Saf*. 2001;10(2):113–9.
25. Levinson D. Adverse events in hospitals: national incidence among medicare beneficiaries. 2010. Department of Health and Human Services. Office of the Inspector General (OEI-06-09-00090).
26. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. *J Gen Intern Med*. 2003;18(1):57–60.
27. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*. 1998;279(15):1200–5.
28. Bond CA, Raehl CL. Adverse drug reactions in United States hospitals. *Pharmacotherapy*. 2006;26(5):601–8.
29. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ*. 2004;329(7456):15–9.
30. Guidance for Industry and FDA Staff-Best Practices for Conducting and Reporting Pharmacoeconomic Safety Studies Using Electronic Healthcare Data U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) U.S. Department of Health and Human Services. 2013. <https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf>. Accessed 10 Aug 2018.
31. Nissen T, Wynn R. The clinical case report: a review of its merits and limitations. *BMC Res Notes*. 2014;7:264.
32. Woodcock J, Behrman RE, Dal Pan GJ. Role of postmarketing surveillance in contemporary medicine. *Annu Rev Med*. 2011;62:1–10.
33. Moore TJ, Cohen MR, Furberg CD. Serious adverse drug events reported to the Food and Drug Administration, 1998–2005. *Arch Intern Med*. 2007;167(16):1752–9.
34. Medication Safety Tools and Practices: ISMP. <https://www.ismp.org/default.asp>. Accessed 10 Aug 2018.
35. AHRQ Topic: Adverse Drug Events Rockville, MD: AHRQ. <http://www.ahrq.gov/health-care-information/topics/topic-adverse-drug-events-ade.html>. Accessed 10 Aug 2018.
36. ASHP Guidelines on Adverse Drug Reaction Monitoring and Reporting: ASHP. <https://www.ashp.org/doclibrary/bestpractices/medmisgdladr.aspx>. Accessed 10 Aug 2018.
37. Trontell A. Current adverse event reporting systems. Adverse drug event reporting: the roles of consumers and health care professionals. Washington, DC: Institute of Medicine of the National Academies Press; 2007. p. 12–7.
38. Inacio P, Cavaco A, Airaksinen M. The value of patient reporting to the pharmacovigilance system: a systematic review. *Br J Clin Pharmacol*. 2017;83(2):227–46.
39. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc*. 2015;22(3):671–81.
40. Pimpalkhute P, Patki A, Nikfarjam A, Gonzalez G. Phonetic spelling filter for keyword selection in drug mention mining from social media. *AMIA Jt Summits Transl Sci Proc*. 2014;2014:90–5.
41. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform*. 2015;53:196–207.
42. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform*. 2017;26(1):214–27.
43. Emadzadeh E, Sarker A, Nikfarjam A, Gonzalez G. Hybrid semantic analysis for mapping adverse drug reaction mentions in tweets to medical terminology. *AMIA Annu Symp Proc*. 2017;2017:679–88.
44. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. *AMIA Annu Symp Proc*. 2014;2014:924–33.
45. Unified Medical Language System: National Institutes of Health. 2013. <https://www.nlm.nih.gov/pubs/factsheets/umls.html>.
46. van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf*. 2002;11(1):3–10.
47. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;18(6):427–36.
48. Molokhia M, Tanna S, Bell D. Improving reporting of adverse drug reactions: systematic review. *Clin Epidemiol*. 2009;1:75–92.
49. Betts KA, Griffith J, Friedman A, Zhou ZY, Signorovitch JE, Ganguli A. An indirect comparison and cost per responder analysis of adalimumab, methotrexate and apremilast in the treatment of methotrexate-naïve patients with psoriatic arthritis. *Curr Med Res Opin*. 2016;32(4):721–9.
50. Billioud V, Sandborn WJ, Peyrin-Biroulet L. Loss of response and need for adalimumab dose intensification in Crohn's disease: a systematic review. *Am J Gastroenterol*. 2011;106(4):674–84.
51. Hochberg MC, Tracy JK, Hawkins-Holt M, Flores RH. Comparison of the efficacy of the tumour necrosis factor alpha blocking agents adalimumab, etanercept, and infliximab when added to methotrexate in patients with active rheumatoid arthritis. *Ann Rheum Dis*. 2003;62(Suppl 2):ii13–6.
52. Kristensen LE, Christensen R, Bliddal H, Geborek P, Danneskiold-Samsøe B, Saxne T. The number needed to treat for adalimumab, etanercept, and infliximab based on ACR50 response in three randomized controlled trials on established rheumatoid arthritis: a systematic literature review. *Scand J Rheumatol*. 2007;36(6):411–7.
53. Paul S, Moreau AC, Del Tedesco E, Rinaudo M, Phelip JM, Genin C, et al. Pharmacokinetics of adalimumab in inflammatory bowel diseases: a systematic review and meta-analysis. *Inflamm Bowel Dis*. 2014;20(7):1288–95.
54. Silva-Ferreira F, Afonso J, Pinto-Lopes P, Magro F. A systematic review on infliximab and adalimumab drug monitoring: levels, clinical outcomes and assays. *Inflamm Bowel Dis*. 2016;22(9):2289–301.
55. Thorlund K, Druyts E, Toor K, Mills EJ. Comparative efficacy of golimumab, infliximab, and adalimumab for moderately to severely active ulcerative colitis: a network meta-analysis accounting for differences in trial designs. *Expert Rev Gastroenterol Hepatol*. 2015;9(5):693–700.
56. Chen X, Hou J, Yuan Y, Huang C, Liu T, Mo C, et al. Adalimumab for moderately to severely active ulcerative colitis: a systematic review and meta-analysis. *BioDrugs*. 2016;30(3):207–17.
57. Eng G, Stoltenberg MB, Szkudlarek M, Bouchelouche PN, Christensen R, Bliddal H, et al. Efficacy of treatment intensification



- with adalimumab, etanercept and infliximab in rheumatoid arthritis: a systematic review of cohort studies with focus on dose. *Semin Arthritis Rheum*. 2013;43(2):144–51.
58. Kawalec P, Mikrut A, Wisniewska N, Pilc A. Tumor necrosis factor- $\alpha$  antibodies (infliximab, adalimumab and certolizumab) in Crohn's disease: systematic review and meta-analysis. *Arch Med Sci*. 2013;9(5):765–79.
  59. Thorlund K, Druyts E, Mills EJ, Fedorak RN, Marshall JK. Adalimumab versus infliximab for the treatment of moderate to severe ulcerative colitis in adult patients naive to anti-TNF therapy: an indirect treatment comparison meta-analysis. *J Crohns Colitis*. 2014;8(7):571–81.
  60. Doan QV, Chiou CF, Dubois RW. Review of eight pharmacoeconomic studies of the value of biologic DMARDs (adalimumab, etanercept, and infliximab) in the management of rheumatoid arthritis. *J Manag Care Pharm*. 2006;12(7):555–69.
  61. Downey C. Serious infection during etanercept, infliximab and adalimumab therapy for rheumatoid arthritis: a literature review. *Int J Rheum Dis*. 2016;19(6):536–50.
  62. Fenix-Caballero S, Alegre-del Rey EJ, Castano-Lara R, Puigventos-Latorre F, Borrero-Rubio JM, Lopez-Vallejo JF. Direct and indirect comparison of the efficacy and safety of adalimumab, etanercept, infliximab and golimumab in psoriatic arthritis. *J Clin Pharm Ther*. 2013;38(4):286–93.
  63. Kopylov U, Al-Taweel T, Yaghoobi M, Nauche B, Bitton A, Lakatos PL, et al. Adalimumab monotherapy versus combination therapy with immunomodulators in patients with Crohn's disease: a systematic review and meta-analysis. *J Crohns Colitis*. 2014;8(12):1632–41.
  64. Machado MA, Maciel AA, de Lemos LL, Costa JO, Kakehasi AM, Andrade EI, et al. Adalimumab in rheumatoid arthritis treatment: a systematic review and meta-analysis of randomized clinical trials. *Rev Bras Reumatol*. 2013;53(5):419–30.
  65. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health Technol Assess*. 2016;20(35):1–610.
  66. Yang Z, Ye XQ, Zhu YZ, Liu Z, Zou Y, Deng Y, et al. Short-term effect and adverse events of adalimumab versus placebo in inducing remission for moderate-to-severe ulcerative colitis: a meta-analysis. *Int J Clin Exp Med*. 2015;8(1):86–93.
  67. Archer R, Tappenden P, Ren S, Martyn-St James M, Harvey R, Basarir H, et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (including a review of TA140 and TA262): clinical effectiveness systematic review and economic model. *Health Technol Assess*. 2016;20(39):1–326.
  68. Burmester GR, Panaccione R, Gordon KB, McIlraith MJ, Lacerda AP. Adalimumab: long-term safety in 23 458 patients from global clinical trials in rheumatoid arthritis, juvenile idiopathic arthritis, ankylosing spondylitis, psoriatic arthritis, psoriasis and Crohn's disease. *Ann Rheum Dis*. 2013;72(4):517–24.
  69. Askling J, Fahrback K, Nordstrom B, Ross S, Schmid CH, Symmons D. Cancer risk with tumor necrosis factor  $\alpha$  (TNF) inhibitors: meta-analysis of randomized controlled trials of adalimumab, etanercept, and infliximab using patient level data. *Pharmacoepidemiol Drug Saf*. 2011;20(2):119–30.
  70. Dretzke J, Edlin R, Round J, Connock M, Hulme C, Cieczot J, et al. A systematic review and economic evaluation of the use of tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) inhibitors, adalimumab and infliximab, for Crohn's disease. *Health Technol Assess*. 2011;15(6):1–244.
  71. Huang ML, Ran ZH, Shen J, Li XB, Xu XT, Xiao SD. Efficacy and safety of adalimumab in Crohn's disease: meta-analysis of placebo-controlled trials. *J Dig Dis*. 2011;12(3):165–72.
  72. Leonardi C, Papp K, Strober B, Reich K, Asahina A, Gu Y, et al. The long-term safety of adalimumab treatment in moderate to severe psoriasis: a comprehensive analysis of all adalimumab exposure in all clinical trials. *Am J Clin Dermatol*. 2011;12(5):321–37.
  73. McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, et al. Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation. *Health Technol Assess*. 2007;11(28):1–158 (iii–iv).
  74. Shepherd J, Cooper K, Harris P, Picot J, Rose M. The clinical effectiveness and cost-effectiveness of abatacept, adalimumab, etanercept and tocilizumab for treating juvenile idiopathic arthritis: a systematic review and economic evaluation. *Health Technol Assess*. 2016;20(34):1–222.
  75. Song YN, Zheng P, Xiao JH, Lu ZJ. Efficacy and safety of adalimumab for the Crohn's disease: a systematic review and meta-analysis of published randomized placebo-controlled trials. *Eur J Clin Pharmacol*. 2014;70(8):907–14.
  76. Wang H, Zuo D, Sun M, Hua Y, Cai Z. Randomized, placebo controlled and double-blind trials of efficacy and safety of adalimumab for treating ankylosing spondylitis: a meta-analysis. *Int J Rheum Dis*. 2014;17(2):142–8.
  77. Wiens A, Correr CJ, Venson R, Otuki MF, Pontarolo R. A systematic review and meta-analysis of the efficacy and safety of adalimumab for treating rheumatoid arthritis. *Rheumatol Int*. 2010;30(8):1063–70.
  78. Chen YF, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, et al. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technol Assess*. 2006;10(42):1–229 (iii–iv, xi–xiii).
  79. Da W, Zhu J, Wang L, Lu Y. Adalimumab for Crohn's disease after infliximab treatment failure: a systematic review. *Eur J Gastroenterol Hepatol*. 2013;25(8):885–91.
  80. Dziechciarz P, Horvath A, Kierkus J. Efficacy and safety of adalimumab for paediatric crohn's disease: a systematic review. *J Crohns Colitis*. 2016;10(10):1237–44.
  81. Ma C, Panaccione R, Heitman SJ, Devlin SM, Ghosh S, Kaplan GG. Systematic review: the short-term and long-term efficacy of adalimumab following discontinuation of infliximab. *Aliment Pharmacol Ther*. 2009;30(10):977–86.
  82. Navarro-Sarabia F, Ariza-Ariza R, Hernandez-Cruz B, Villanueva I. Adalimumab for treating rheumatoid arthritis. *Cochrane Database Syst Rev*. 2005;3:CD005113.
  83. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technol Assess*. 2011;15(10):1–329 (i–xxi).
  84. Diaz-Lagares C, Belenguer R, Ramos-Casals M. Systematic review on the use of adalimumab in autoimmune. Efficacy and safety in 54 patients. *Reumatol Clin*. 2010;6(3):121–7.
  85. Malottki K, Barton P, Tsourapas A, Uthman AO, Liu Z, Routh K, et al. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. *Health Technol Assess*. 2011;15(14):1–278.
  86. Shepherd J, Vidt DG, Miller E, Harris S, Blasetto J. Safety of rosuvastatin: update on 16,876 rosuvastatin-treated patients in a multinational clinical trial program. *Cardiology*. 2007;107(4):433–43.
  87. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignat J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. *J Med Internet Res*. 2015;17(7):e171.

88. Bate A, Lindquist M, Edwards IR, Orre R. A data mining approach for signal detection and analysis. *Drug Saf.* 2002;25(6):393–7.
89. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *JAMA.* 1997;277(4):301–6.
90. Pages A, Bondon-Guitton E, Montastruc JL, Bagheri H. Undesirable effects related to oral antineoplastic drugs: comparison between patients' internet narratives and a national pharmacovigilance database. *Drug Saf.* 2014;37(8):629–37.
91. Pierce CE, Bourri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of facebook and twitter monitoring to detect safety signals for medical products: an analysis of recent fda safety alerts. *Drug Saf.* 2017;40(4):317–31.
92. Bilimoria KY, Kmiecik TE, DaRosa DA, Halverson A, Eskandari MK, Bell RH Jr, et al. Development of an online morbidity, mortality, and near-miss reporting system to identify patterns of adverse events in surgical patients. *Arch Surg.* 2009;144(4):305–11 (**discussion 11**).
93. Hinrichsen VL, Kruskal B, O'Brien MA, Lieu TA, Platt R. Using electronic medical records to enhance detection and reporting of vaccine adverse events. *J Am Med Inform Assoc.* 2007;14(6):731–5.